UNIVERSITY OF GOTHENBURG
SCHOOL OF BUSINESS, ECONOMICS AND LAW

WORKING PAPERS IN ECONOMICS

No 649

How are you? How's it going? What's up? What's happening? Nudging people to tell us how they really are

Fredrik Carlsson and Mitesh Kataria

March 2016

ISSN 1403-2473 (print)
ISSN 1403-2465 (online)

Department of Economics
School of Business, Economics and Law at University of Gothenburg
Vasagatan 1, PO Box 640, SE 405 30 Göteborg, Sweden
+46 31 786 0000, +46 31 786 1326 (fax)
www.handels.gu.se info@handels.gu.se

EFMD EQUIS ACCREDITED

How are you? How's it going? What's up? What's happening?
Nudging people to tell us how they really are

Fredrik Carlsson,[a] Mitesh Kataria[b]

Abstract
We investigate a novel approach to reduce measurement error in subjective well-being (SWB) data. Using a between-subject design, half of the subjects are asked to promise to answer the survey questions truthfully in an attempt to make them commit to truth-telling. This allows us to experimentally test whether making a promise affects their responses. We find a statistically significant difference between mean stated well-being between the two groups (with and without a promise, although the effect sizes are rather small). We then investigate to what extent the differences in stated well-being also affect the inference from regressions models on the determinants of SWB. We find important differences in terms of size and statistical significance of the coefficients between the two models, despite the small effect sizes on the dependent stated well-being variable.

Key-words: measurement error, social desirability, subjective well-being, truth-telling

JEL-classifications: C90, I30,

[a] Department of Economics, University of Gothenburg, Box 640, SE-40530 Gothenburg, Sweden. fredrik.carlsson@economics.gu.se
[b] Department of Economics, University of Gothenburg, Box 640, SE-40530 Gothenburg, Sweden. mitesh.kataria@economics.gu.se

1. Introduction

We have witnessed an increased use of subjective well-being (SWB) measures in economics; from 2000 to 2006, 157 papers and numerous books on the topic were published in the economics literature (Krueger and Schkade, 2008). While most economists would probably agree that the information gained from subjective questions is interesting and important, the unwillingness to rely on such questions has historically marked an important difference between economists and other social scientists (Bertrand and Mullainathan, 2001). This attitude may, however, have shifted among economists during the last ten years. Furthermore, we have seen an increased interest in incorporating findings from other disciplines, such as psychology, into economics (see, e.g., Layard, 2006). One of the problems with SWB data is that it is prone to social desirability bias, i.e., the tendency of survey respondents to answer questions in a manner that will be viewed favorably by others and in line with certain social norms (Phillips and Clancy, 1972). Such social-image concerns are likely to be important to many people (Lacetera and Macis, 2010). Although not as extensively explored, it is also plausible that self-image concerns can bias survey responses, e.g., in an SWB context respondents may not want to admit even to themselves that they are unsatisfied with their life or some aspects thereof. These types of misreportings create a measurement error in the reported data, which can be handled either at the modeling stage[1], the data collection stage, or both.

In this paper we investigate a novel approach to reduce measurement error in SWB surveys. We test a self-commitment mechanism where survey respondents are asked to promise to answer the survey questions truthfully. The main objective is to experimentally test whether making a promise affects responses to survey questions, in particular subjective well-being questions. Subsequently, if we find differences, we investigate to what extent these differences affect the inference drawn from regressions models on the determinants of SWB.

Traditionally, economists have been skeptical to asking respondents to tell the truth, primarily because there are no actual incentives for responding truthfully. However, empirical evidence suggests that a promise alone can indeed induce an emotional commitment to fulfill the promise (Braver, 1995; Ostrom et al., 1992; Ellingsen and Johannesson, 2004; Vanberg 2008; Carlsson et al. 2013; Kataria and Winter, 2013; Jacquemet et al., 2013).[2] This approach

[1] See, e.g., Hausman (2001) for a review. The mismeasurement problem for linear models in econometrics is usually solved by using instrumental variables (Hausman, 2001). Abrevaya and Hausman (1999) offer methods that also deal with binary choice with misclassification and mismeasured discrete dependent variables with several categories.

[2] The reason why it works has recently been under scientific scrutiny. Charness and Dufwenberg (2006) found experimental evidence that a promise works because of guilt aversion: A guilt-averse person does not want to let

has been applied in, among other areas, experiments (Ellingsen and Johannesson, 2004; Jacquemet et al., 2013) and stated preference surveys on environmental problems (Carlsson et al., 2013). Yet as far as we know, it has not been used in subjective well-being surveys.

The literature on how to reduce measurement error at the data collection stage is rather extensive. One such area is survey research dealing with particularly sensitive topics such as racism, terrorism, corruption, illegal behavior, and drug use. Methods such as randomized response (Warner 1965; Greenberg 1969) and item count techniques (Raghavarao and Federer, 1979) have been used in surveys with sensitive questions. Using randomized response, the respondent flips a coin and is instructed to answer either a sensitive or a non-sensitive yes/no question based on the outcome of the coin flip. Only the respondent knows which question he or she answered. This procedure hides individual answers but enables analysts to assess the true population proportions of yes/no replies because the noise probability is known. The drawback of the randomized response method is that it draws attention to the act of measurement itself. Respondents can become suspicious of intent and claims to anonymity and focus too much on how the method works instead of answering the questions. The item count method randomly splits the sample into two groups: control and treatment group. Both groups are asked insensitive questions, and then the treatment group is also asked a sensitive question, i.e. the question of interest. The respondents are then asked to reveal the number of "yes" answers they have given. Respondent anonymity is assured whereas the number of people who answered "yes" to the sensitive question can be mathematically deduced. While the item count method is straightforward with a low level of burden on the respondents, one major drawback of the method is the lack of power of the estimator for relatively high sample sizes (Droitcour, 1991).

SWB data utilizes what in economics is known as experience utility, which is distinguished from what is known as decision utility. While decision utility is inferred from the decision maker’s observed choices, experience utility is the satisfaction we experience once the decision is made. If the choices are revealed in a market, the data is known as revealed preference data while stated preference data represents choices made or stated in a constructed survey situation. Stated preference data is used frequently in economics to value public goods. Interestingly, both SWB data and stated-preference studies deal with the same difficulties common to data based on subjective assessments. A number of approaches to

down others’ expectations and is therefore committed to the promises made. An alternative explanation is that people may have a taste for keeping their word (e.g., Ellingsen and Johannesson, 2004). Using a novel design, Vanberg (2008) found support for the latter explanation, i.e., people have preferences for promise keeping per se.

reduce the so-called hypothetical bias, i.e., the difference between stated behavior and behavior if the choice situation would have been an actual one, have been suggested in the stated preference literature. One common approach is to use a cheap talk script, which aims to reduce hypothetical bias by informing respondents about the occurrence of hypothetical bias (Cummings and Taylor, 1999). The idea is that by making respondents aware of the problem, they will exert more effort when responding and in that way hypothetical bias will be reduced. The empirical support for a cheap talk script is mixed, and it is clear that the effect depends both on the context and on the formulation of the script (see, e.g., Aadland and Caplan, 2006; Carlsson et al., 2005; List, 2001). Another approach that has been used is so-called inferred valuation (Carlsson et al., 2010; Lusk and Norwood, 2009a,b), where respondents estimate other people's valuations of goods.[3] Compared with the stated preference literature, surprisingly little attention has been given to reduce measurement error in SWB surveys. Layard (2006) argues that there is a need for an expanded model of happiness that incorporates findings from other disciplines, such as psychology. While his main focus was on theory, there is without doubt also a need to develop the measurement methods of SWB data, and as discussed, it seems possible to incorporate findings from neighboring disciplines in the SWB literature.

Understanding what affects subjective well-being is important as it could help economists design policies that improve people's wellbeing. In the late 1990s, economists started to present large-scale empirical analyses of determinants of well-being (Frey and Stutzer, 2002). Evidence suggests that poor health, divorce, unemployment, and lack of social relationships are important determinants of well-being (Dolan et al., 2008). While economists usually focus on determinants of SWB that can be categorized as actual observable life experiences, the approach has been somewhat more pluralistic in psychology. Here, the overall well-being is complemented with reported well-being in major domains of life, such as health, finances, social relationships, and sex life. The advantage of this life domain approach is that it better reflects subjective factors such as personal aspirations and norms that could affect the overall well-being. For example, an individual with high income and high financial aspirations could be less satisfied with life than someone with low income and low financial aspirations. In this paper we apply both approaches, i.e. both the determinants of overall and domain-specific SWB. Since social desirability and self-image bias is most likely a problem for value-laden

[3] Other approaches to reduce hypothetical bias in the stated preference literature include ex-post calibration of willingness-to-pay estimates based on follow-up questions (e.g., Champ et al., 1997; Champ and Bishop, 2001; Johannesson et al., 1999) and the time-to-think protocol (Cook et al., 2007; Whittington et al., 1992).

questions, it is of interest to investigate the potential effect of measurement errors both in a setting with only a measurement error in the dependent variable and in a setting with measurement errors in both the dependent and some independent variables.

2. Social desirability and self-image in a measurement-error framework

One of the problems with SWB data is that it is prone to social desirability bias, i.e., the tendency of survey respondents to answer questions in a manner that will be viewed favorably by others and in line with some social norm (Phillips and Clancy, 1972). Self-image concerns may also bias the survey responses, e.g., respondents do not want to admit even to themselves that they are not satisfied with their life.

In order to illustrate the potential problem with social desirability and self-image concerns, we use a measurement-error framework where the dependent and/or the independent variables in a regression model are observed with an error. The simplest case is a linear regression model with one independent variable and no intercept, which is also the standard textbook case (see, e.g., Greene, 2002). The observed dependent variable, y, is specified as $y = y^* + u$, where y^* is the true variable and u is a normally distributed error term, i.e., $u \sim N(\mu_y, {\sigma^2}_y)$. Suppose that the observed independent variable is x, i.e. $x = x^* + v$, where x^* is the true variable and $v \sim N(\mu_x, \sigma_x)$. Assume that u and v are independently distributed. If $\mu = 0$ the error term is a random error, and if $\mu \neq 0$ the error is a systematic error. Two results are well-known in the literature. First, assuming that only y is measured with a random error does not result in biased parameter estimates since the measurement error is incorporated in the disturbance term. It will, however, increase the standard error of the estimated parameter, i.e., the parameter will be estimated with less precision. Second, if instead x is measured with a random error, the parameter estimates are inconsistent and biased towards zero (attenuation bias). Hausman (2001) calls this the “iron law of econometrics” – the magnitude of the estimate is usually smaller than expected. If both the dependent and the independent variable are mismeasured, the parameters are still – of course – biased and measured with less accuracy. The regression models used in the SWB literature deviate in many aspects from the simplifying assumptions made above. First of all, a multiple regression framework, with an intercept, is used in the litterature. Second, if social desirability and self-image are the reasons for measurement error, we would expect the dependent variable to be measured with a systematic rather than a random error. This is also true for independent variables with value-laden content such as life satisfaction in different domains, but perhaps not for objective variables that are merely counts of various types of individual characteristics. With objective variables we expect the problem of measurement error to be less severe, and if present we would expect it to be a random error. Also note that the measurement errors of the independent variables might be correlated with each other and the measurement error of the dependent variable. Third, the dependent variable is measured on an ordinal scale and should therefore be estimated with a non-linear model such as an ordered probit model. Hausman et al. (1998) showed that misclassification[4] of the dependent discrete variable causes inconsistent coefficient estimates if the measurement error is not taken into consideration in a standard framework (e.g., probit or logit). Relatively small amounts of misclassification of the dependent variable can lead to a large bias even with a large sample size. Hence, measurement bias causes severe problems. Exactly how these problems will manifest in our application is an empirical question. We will address the issue of measurement error in SWB data experimentally by comparing stated levels of well-being and coefficients of regressions models from two different survey versions, where only one of the versions include a short script and question asking if the respondents can answer the questions in the survey truthfully or not.

3. Survey design

The questions used in the present paper were administered to respondents as part of the thirteenth wave of the Citizen Panel (Martinsson et al., 2014). The Citizen Panel is an online panel survey administered by the Laboratory of Opinion Research (LORE), which was established in 2010 by the Multidisciplinary Opinion and Democracy (MOD) research group at the Faculty of Social Science, University of Gothenburg in Sweden. The survey was carried out from November 27 to December 21 in 2014, and consisted of a set of core questions that were combined with some specific questions for the purpose of this study asked at the end of the survey. The Citizen Panel consists mainly of self-recruited respondents (85 percent). The remaining respondents (15 percent) is a probability-based recruitment from population samples. Overall, we consider the data to be of sufficient quality for the purpose of this paper. However, there are of course reasons to be cautious, especially when looking beyond the differences between the two treatments in an attempt to interpret what affects subjective well-being.

[4] Misclassification means that the response is reported or recorded in the wrong category.

The main feature of the experiment was that, based on random allocation, the subjects received either a survey version asking them to promise to tell the truth or a version without such a request. In all other respects, the two survey versions were identical.

Immediately after the request to promise to tell the truth, the survey consisted of questions about overall and domain-specific stated well-being. More specifically, the respondents were asked how satisfied they felt overall with their life, and subsequently with various aspects of their life on a scale from 0 to 10. Finally they were asked questions about social trust, social interaction, health status, and socio-economic characteristics.

4. Results

5.1. Descriptive results

Table 1 reports descriptive statistics for the responses to the general and domain-specific well-being questions. The last columns report a test of the difference between the mean values with and without a promise, as well as the effect size.

Table 1. Overall and domain-specific stated well-being, descriptive statistics, and test of difference between No Promise and Promise

	No promise			Promise			Difference		Effect size
	Mean	*St.Dev.*	*Obs.*	*Mean*	*St.Dev.*	*Obs.*	*Mean*	*P-value*	*Cohen's d*
Overall	6.94	1.99	1,782	6.58	2.10	1,700	0.36	0.000	0.176
Financial situation	6.06	2.53	1,777	5.92	2.61	1,699	0.14	0.057	0.056
Spare time	6.55	2.15	1,761	6.25	2.22	1,688	0.30	0.000	0.140
Work	6.12	2.55	1,506	5.97	2.64	1,427	0.15	0.064	0.034
Social life	6.88	2.18	1,769	6.52	2.25	1,695	0.36	0.000	0.164
Sex life	5.30	2.94	1,648	4.83	2.99	1,629	0.47	0.000	0.188
Family life	7.13	2.37	1,669	6.95	2.43	1,610	0.18	0.016	0.078
Relationships	7.35	2.59	1,401	7.05	2.73	1,335	0.30	0.003	0.076
Health	6.27	2.61	1,769	6.08	2.61	1,698	0.19	0.027	0.075

* All variables range from 0 to 10.

As discussed, the expected effect of the promise treatment is a reduction in stated well-being. This is confirmed for all nine well-being measures. Moreover, using a one-sided t-test we find that the differences are statistically significant at a 5 percent significance level apart from financial situation and work, which are significant at a 10 percent level. The difference in mean values is largest for the sex life domain. The largest effect size is found for the overall well-being measure, with a Cohen's d of 0.18. Consequently, although there is a robust effect of the promise treatment, the effect sizes are rather small.[5] However, as discussed above, a relatively small amount of misclassification of a discrete dependent variable can still lead to biased coefficient estimates in the regression analysis (Hausman et al., 1998).

We also report descriptive statistics of the variables that will be used in the regression models. These include both objective variables that we do not expect to be affected by the promise treatment, and a set of more value-laden questions such as self-reported health. The variables are presented in Table 2.

Table 2. Potential determinants of stated well-being, descriptive statistics, and test of difference between the treatments (with a promise and the control group without a promise)

Variable	Description	No promise		Promise		Difference
		Mean	*Std dev*	*Mean*	*Std dev*	*P-value (t/pr-test)*
Woman	= 1 if subject is a woman	0.47		0.47		0.772
Age	Age in years	48.1	14.2	47.6	14.1	0.319
No of children	No. of children (< 18 years) living in household	1.42	1.28	1.40	1.28	0.547
Divorced	= 1 if divorced	0.07		0.07		0.589
Unemployed	= 1 if unemployed	0.03		0.04		0.463
University	= 1 if university education (at least 3 years)	0.32		0.32		0.927
Income	Individual monthly income before taxes in SEK	30 900	15 350	30 300	15 100	0.261
Reported health	Self-reported health status, 1= very poor; 5 = very good	3.71	0.98	3.65	1.00	0.075
Social trust	Stated trust, 1 = low trust;10 = high trust	6.61	2.20	6.51	2.23	0.179

[5] With a Cohen's d of 0.2, 58 % of the treatment group would be above the mean of the control group, 92 % of the two groups would overlap, and there would be a 56 % chance that a person randomly picked from the treatment group would have a higher score than a person randomly picked from the control group, i.e., 0.56 is the probability of superiority (McGraw and Wong, 1992).

Social interact. Low	= 1 if interact with friends/ relatives less than once per month	0.10	0.11	0.417
Social interact. intermediate	= 1 if interact with friends/ relatives at least once per month	0.42	0.43	0.882
Social interact. High	= 1 if interact with friends/ relatives at least once per week	0.47	0.46	0.602

As expected, there are no statistically significant differences among the objective variables between the versions with and without a promise. However, for the more subjective question about self-reported health status, we do find a statistically significant difference. Self-reported health is higher without a promise. This is in line with Jurges (2006), who found that Danish and Swedish respondents tend to largely overrate their health. No statistically significant differences between the versions with and without a promise are found for the other questions of a more of subjective nature, such as social trust and social interaction.

5.2 Regression models

So far we have confirmed a systematic effect on subjective well-being by asking subjects to promise to tell the truth. In addition, we found a statistically significant difference in self-reported health status between the two survey versions, while for the other subjective and all objective questions we did not find any statistically significant differences. The next question is whether the differences between the two survey versions affect coefficient estimates – in terms of size and statistical significance – in regression models of SWB. In order to investigate this, we compare two regression models with the same model specification that only differ in whether data was collected using a survey with or without asking the respondents to promise to tell the truth. Since the data is ordinal, we estimate ordered probit models.[6] The results are presented in Table 3.

Table 3. Ordered probit models, stated SWB as dependent variable, with and without a promise to tell the truth

	No Promise	**Promise**	**Difference**
	Coeff (S.E.)	*Coeff (S.E.)*	*P-value Chi-2 test*

[6] As discussed by Ferrer-Carbonell and Frijters (2004), the empirical findings in SWB studies need not be sensitive to the choice between a standard OLS model and a discrete model such as an ordered probit model. However, in our specific case we focus on a model that from a conceptual point of view is the more appropriate model because measurement errors are more problematic in a discrete model framework.

Woman	0.156***	0.153***	0.971
	(0.053)	(0.053)	
Age	0.006***	0.008***	0.474
	(0.002)	(0.002)	
No. of children	0.138***	0.127***	0.746
	(0.023)	(0.023)	
Divorced	-0.337***	-0.589***	0.091
	(0.101)	(0.104)	
Unemployed	-0.883***	-0.499***	0.098
	(0.148)	(0.147)	
University	-0.154***	-0.108*	0.564
	(0.057)	(0.058)	
Income	0.020	0.067***	0.078
	(0.017)	(0.018)	
Reported health	0.522***	0.477***	0.312
	(0.029)	(0.028)	
Social trust	0.089***	0.097***	0.659
	(0.089)	(0.012)	
Social interaction low	-0.137	-0.380***	0.070
	(0.089)	(0.088)	
Social interaction high	0.143***	0.152***	0.905
	(0.054)	(0.055)	
No. obs.	1659	1596	
Pseudo R2	0.097	0.094	

*, **, and *** denote significance at 10%, 5%, and 1%, respectively.

In the model based on data from the survey without a promise, we see that most coefficients are statistically significant and the signs are in line with what is typically found in SWB studies. Stated well-being is positively correlated with being a woman, age, number of children, health status, and social trust, and negatively correlated with being divorced, unemployed, and a low level of social interaction. The exceptions are that individual income does not have a statistically significant effect on SWB and that university education has a statistically significant negative impact on SWB. The results are largely similar in terms of sign and statistical significance in the model based on data with a promise to tell the truth. However, at a the 10 percent significance level there are four significant differences between the two datasets, which is evidence that a measurement bias can lead to incorrect inferences in the regression analysis of SWB data. The estimated negative effect of being divorced is greater when using the data from the survey with a promise to tell the truth. In contrast, the estimated negative effect of being unemployed is smaller when using the data with a promise. Moreover, with a promise we observe that individual income has a positive and statistically significant effect on subjective well-being. Note that the reported incomes and proportions of subjects who are divorced or unemployed do not differ between the two survey versions. Finally, the negative impact of having limited social interaction is only statistically significant in the version with a promise, and the difference between the two versions is statistically significant.

In previous studies it has been found that one effect of asking people to promise to tell the truth is that the underlying variance decreases (see, e.g., Carlsson et al., 2013). In order to control for differences in variance between the treatments for SWB data, we estimate a heteroskedastic ordered probit model on the pooled data. We include a set of interaction terms for all independent variables in order to allow for a level difference between the two survey versions as well. The results are presented in Table A1 in the appendix. The differences between the two survey versions remain the same as when comparing the two models in Table 3, and there is no statistically significant difference in variance between the two survey versions. Thus, in this specific case, there is no effect on the underlying variance when asking subjects to promise to tell truth.

So far we have some evidence that a measurement bias can lead to incorrect inference in the regression analysis of SWB data. The results are in line with the prediction of attenuation bias since the absolute value of the coefficients are smaller in the version without a promise for three of the four coefficients where we observe a statistically significant difference between the two survey versions.

Finally, we investigate the relationship between the stated overall well-being and the stated well-being in specific domains. We do this by estimating a model with the domain-specific well-being as independent variables. The main focus is to compare the results between the two survey versions with and without a promise to tell the truth.

Table 4: Ordered probit models, stated overall SWB as dependent variable, and well-being in specific domains as independent variables, with and without a promise to tell the truth

	No Promise	Promise	Difference
	Coeff (S.E.)	Coeff (S.E.)	P-value Chi-2 test
Financial situation	0.110	0.121	0.635
	(0.016)	(0.015)	
Spare time	0.122	0.159	0.264
	(0.019)	(0.019)	
Work	0.154	0.151	0.895
	(0.015)	(0.015)	
Social life	0.048	0.037	0.735
	(0.019)	(0.019)	
Sex life	0.003	0.002	0.952
	(0.014)	(0.014)	
Family life	0.119	0.127	0.823
	(0.021)	(0.020)	
Relationships	0.079	0.070	0.778
	(0.019)	(0.018)	
Health	0.101	0.105	0.867
	(0.015)	(0.014)	
No. obs.	1178	1115	
Pseudo R2	0.229	0.233	

All coefficients are significant at the 1 percent level.

Are there any significant differences in the estimated coefficients between the two regression models? Surprisingly, although Table 3 revealed that people significantly overstate their overall as well as domain-specific satisfaction with life, there are no notable differences in the estimated coefficients comparing the models with and without a promise. Actually, the implicit ranking of the domains based on the size of the coefficient estimates is exactly the same in the two models. More importantly, the last column in Table 4 also shows that there are no statistically significant differences. Thus, the estimated correlations between domain-specific and overall well-being are not affected by a promise to tell the truth. This is true even for the presumably sensitive issue sex life.

In order to investigate whether the results to any extent depend on a potential effect on the underlying variance, we estimate a heteroskedastic ordered probit model as well. Results are presented in Table A2 in the appendix. Although the statistical significance of a few coefficients is affected, the overall conclusion is unaffected. Nor do we find any statistically significant effect on the underling variance of the promise treatment.

5. Conclusions

Social-desirability and self-image concerns are factors that might affect how subjects respond to survey questions, in particular value-laden questions about for example how happy you are or how satisfied you are with your life. We find a systematic effect on stated well-being of making the subjects promise to tell the truth. We find this for both overall well-being and subjective well-being in a number of specific domains. In general, we find that people are inclined to overstate their life satisfaction. The effect size for stated subjective well-being is certainly not large; in fact it is rather small. However, even small differences can in theory have important implications for the inferences we draw in our regression models. This is indeed what we find empirically. There are important differences in terms of coefficient sizes and statistical significance between the two models with and without a promise to tell the truth.

One obvious objection to all methods attempting to reduce social desirability, hypothetical bias, and similar effects at the data collection stage is that they are prone to experimenter demand effects. While this is generally true, we believe that it is of less concern in our setting since we do not say anything in the survey about the expected direction of a bias, something that is often done in for example stated preference surveys on public goods. Thus, it is hard for the subjects to know which researcher expectations to comply with beyond the simple request to answer truthfully.

To our knowledge, we are the first to test how promises to answer truthfully affect SWB data. Clearly, more evidence is needed to draw robust conclusions. We also see a need for more work in general to improve the measurement of SWB data.

Acknowledgements

We are very grateful to Elina Lampi for generously sharing her expertise and collaborating with us at the design stage of this study. We also wish to thank the Laboratory of Opinion Research at the University of Gothenburg for supporting this project and for helping us administer the survey and collect the data.

References

Aadland, D., Caplan, A., 2006. Cheap talk revisited: New evidence from CVM. *Journal of Economic Behavior and Organization* 60, 562-578.

Bertrand, M., Mullainathan, S., 2001. Do People Mean What They Say? Implications for Subjective Survey Data. *American Economic Review* 91, 67-72.

Braver, S. L. 1995. Social contracts and the provision of public goods. In D. A. Schroeder (Ed.), Social dilemmas: Perspectives on individuals and groups (pp. 69-86). Westport, CT: Praeger.

Carlsson, F., Frykblom, P., Lagerkvist, C.J., 2005. Using cheap-talk as a test of validity in choice experiments. *Economic Letters* 89, 147-152.

Carlsson F., Daruvala D., Jaldell, H., 2010. Do you do what you say or do you do what you say others do? *Journal of Choice Modeling* 3, 113-133

Carlsson, F., Kataria, M., Krupnick, A., Lampi, E., Löfgren, A., Qin, P., Sterner, T., Chung, S., 2013. The truth, the whole truth, and nothing but the truth: a multiple country test of an oath script. *Journal of Economic Behavior and Organisation 89, 105-121.*

Champ, P., Bishop, R., 2001. Donation Payment Mechanisms and Contingent Valuation: An Empirical Study of Hypothetical Bias. *Environmental and Resource Economics* 19, 383-402

Champ, P.,Bishop, R., Brown, T., McCollum, D., 1997. Using Donation Mechanisms to Value Nonuse Benefits From Public Goods. *Journal of Environmental Economics and Management* 33, 151-162.

Charness, G., Dufwenberg, M., 2006. Promises and partnership. *Econometrica* 74, 1579–1601.

Cook, J., Whittington, D., Canh, D., Johnson, F.R., Nyamete, A., 2007. Reliability of Stated Preferences for Cholera and Typhoid Vaccines with Time to Think in Hue Vietnam. *Economic Inquiry* 45, 100-114.

Cummings, R., Taylor, L., 1999. Unbiased value estimates for environmental goods: a cheap talk design for the contingent valuation method. *American Economic Review* 89, 649-65

Dolan, P., Peasgood, T., White, M., 2008. Do we really know what makes us happy? A review of the economic literature on the factors associated with subjective well-being. *Journal of Economic Psychology,* 29, 94–122.

Droitcour, J., Caspar, R., Hubbard, M., Parsely, T., Vischer, W., Ezzati, T., 1991. The item count technique as a method of indirect questioning: A review of its development and a case study application. In Biemer, P., Groves, R. Lyberg. L., Mathiowwetz, N., Sudman, S. (eds.) *Measurement errors in survey*s, Wiley, New York.

Ellingsen, T., Johannessson, M., 2004. Promises, threats and fairness. *Economic Journal* 114, 397–420.

Ferrir-i-Carbonell, A., Frijters, P., 2004. How important is methodology for the estimates of the determinant of happiness? *Economic Journal* 114, 641-659.

Frey, B., Stutzer A., 2002. What Can Economists Learn from Happiness Research?, *Journal of Economic Literature*, 40, 402-35.

Greenberg, B., Abul-Ela, A-L., Simmons, W., Horvitz, D., 1969. The Unrelated Question Randomized Response Model: Theoretical Framework. Journal of the American Statistical Association 64, 520–39.

Greene, W., 2003. *Econometric analysis* (5th ed.), New Jersey: Prentice Hall.

Hausman, J., 2001. Mismeasured variables in econometric analysis: problems from the right and problems from the left. *The Journal of Economic Perspectives* 15, 57–67.

Hausman, J., Abrevaya, J., Scott-Morton F-M., 1998. Misclassification of the Dependent Variable in a Discrete-Response Setting. *Journal of Econometrics,* 87, 239–69.

Jacquemet, N., Joule, R.V., Luchini, S., Shogren, J. 2011. Preference elicitation under oath. *Journal of Environmental Economics and Management* 48, 413-433.

Johannesson M., Blomquist, G., Blumenschein, K., Johansson, P-O., Liljas, B., O´Conor, R.M., 1999. Calibrating Hypothetical Willingness to Pay Responses, *Journal of Risk and Uncertainty*, 8, 21-32.

Jurges, H. 2007. True health vs. response styles: Exploring cross-country differences in self-reported health. *Health Economics*, 16, 163–178.

Lacetera, N., Macis, M., 2008. Social image concerns and prosocial behavior: Field evidence from a nonlinear incentive scheme. *Journal of Economic Behavior and Organization*, 76, 225–237.

Kataria, M., Winter F., 2013. Third Party Assessments in Trust Problems with Conflict of Interest: An Experiment on the Effects of Promises, *Economics Letters*, 20, 53-56.

Krueger A. B., Schkade, D., 2008. The Reliability of Subjective Well-Being Measures. *Journal of Public Economics* 92,1833-45.

Layard, R., 2006. Happiness and public policy: A challenge to the profession. *The Economic Journal* 116, C24-C33.

List, J., 2001. Do explicit warnings eliminate the hypothetical bias in elicitation procedures? Evidence from field auction experiments. *American Economic Review* 91, 1498–150.

Lusk J., Norwood, B., 2009a. An inferred valuation method. *Land Economics* 85, 500-514.

Lusk J., Norwood B., 2009b. Bridging the gap between laboratory experiments and naturally occurring markets: An inferred valuation method. *Journal of Environmental Economics and Management* 58, 236-250.

McGraw, K. O., Wong, S. P. (1992). A common language effect size statistic. *Psychological Bulletin, 111,* 361–365.

Martinsson, J., Andreasson, M., Markstedt, E., Riedel, K., Samanni, M., 2014. Technical report citizen panel 13 – 2014. LORE, University of Gothenburg.

Ostrom, E., Walker, J., Gardner, R., 1992. Covenants with and without the sword: self-governance is possible. *American Political Science Review* 86, 404–417.

Phillips, D. L., Clancy, K.J., 1972. Some Effects of "Social Desirability" in Survey Studies. *American Journal of Sociology*, 77, 921-940

Raghavarao, D., Federer., W.T., 1979. Block Total Response as an Alternative to the Randomized Response Method in Surveys. *Journal of the Royal Statistical Society, Series* B 41, 40–45.

Vanberg, C., 2008. Why do people keep their promises? An experimental test of two explanations. *Econometrica* 76, 1467–1480.

Warner, S. L. 1965. Randomized response: a survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association* 60 (309): 63–69.

Whittington, D., Smith, V.K., Okorafor, A., Okore, A., Liu, J., McPhail, A., 1992. Giving Respondents Time to Think in Contingent Valuation Studies: A Developing Country Application. *Journal of Environmental Economics and Management* 22, 205–25.

Appendix

Table A1. Heteroskedastic ordered probit model, stated SWB as dependent variable, with and without a promise to tell the truth

	Coeff (S.E.)
Woman	0.157*** (0.053)
Age	0.006*** (0.002)
No. of children	0.139*** (0.023)
Divorced	-0.335*** (0.101)
Unemployed	-0.897*** (0.148)
University	-0.154*** (0.057)
Income	0.021 (0.017)
Reported health	0.524*** (0.029)
Social trust	0.089*** (0.012)
Social interaction low	-0.137 (0.089)
Social interaction high	0.141*** (0.054)
Promise	-0.390*** (0.223)
Woman × Promise	0.003 (0.076)
Age × Promise	0.003 (0.003)
No. of children × Promise	-0.008 (0.032)
Divorce × Promise	-0.273* (0.148)
Unemployed × Promise	0.387* (0.211)
University × Promise	0.042 (0.082)
Income × Promise	0.049* (0.025)
Health × Promise	-0.033 (0.039)
Social trust × Promise	0.012 (0.018)
Social interaction low × Promise	-0.249* (0.128)
Social interaction high × Promise	0.017 (0.079)
Variance	
Promise	0.035 (0.026)
No. obs.	1659

Pseudo R2	0.097

*, **, and *** denote significance at 10%, 5%, and 1%, respectively.

Table A2. Heteroskedastic ordered probit models, stated SWB as dependent variable, and well-being in specific domains as independent variables interacted with and without a promise to tell the truth

	Coeff (S.E.)
Financial situation	0.112*** (0.015)
Spare time	0.121*** (0.019)
Work	0.155*** (0.015)
Social life	0.048** (0.019)
Sex life	0.002 (0.014)
Family life	0.121*** (0.021)
Relationships	0.079*** (0.019)
Health	0.102*** (0.0145)
Promise	-0.351* (0.181)
Financial situation × Promise	0.009 (0.021)
Spare time × Promise	0.038 (0.027)
Work × Promise	-0.003 (0.021)
Social-life × Promise	-0.011 (0.027)
Sex-life × Promise	0.002 (0.020)
Family-life × Promise	0.007 (0.029)
Relationships × Promise	-0.009 (0.026)
Health × Promise	0.003 (0.020)
Variance	
Promise	0.008 (0.033)
No. obs.	2293
Pseudo R2	0.233

*, **, and *** denote significance at 10%, 5%, and 1%, respectively.